

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-157273

(43)Date of publication of application : 30.05.2003

(51)Int.Cl.

G06F 17/30

(21)Application number : 2001-353554

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

(22)Date of filing : 19.11.2001

(72)Inventor : MAKINO TOSHIAKI
SUGIZAKI MASAYUKI
INAGAKI HIROTO

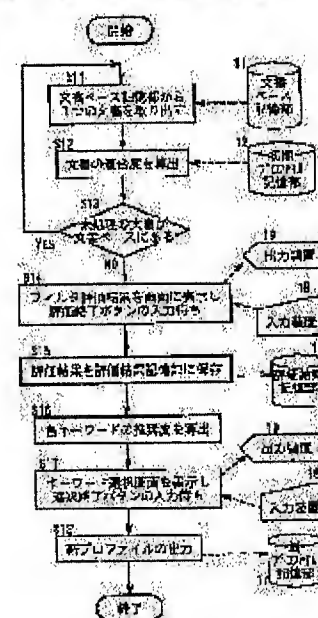
(54) METHOD AND DEVICE FOR AIDING PREPARATION OF PROFILE, RECORDING MEDIUM AND SOFTWARE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a profile preparation aiding method and a device therefor, a recording medium and software capable of easily preparing a suitable profile.

SOLUTION: In the profile preparation aiding method for aiding the preparation of a profile expressing user's interest, filtering for selecting a specific document matching with the user's interest from a plurality of documents is performed by using an initial profiling, the filtering result is outputted to the user, an input corresponding to the result is inputted from the user as evaluation result information expressing a correct answer or incorrect answer, and the contents of a document of the correct answer or the contents of a document of the incorrect answer are analyzed on the basis of the evaluation result. Then the analyzed result is compared with the information of the initial profile, an adding candidate keyword and a deleting candidate keyword are extracted and outputted to the user and the profile is corrected in accordance with a user's selection input of these candidate keywords.

実施の形態のプロファイル作成支援装置の動作



(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2003-157273

(P2003-157273A)

(43)公開日 平成15年5月30日(2003.5.30)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード(参考)
G 0 6 F 17/30	3 4 0	G 0 6 F 17/30	3 4 0 A 5 B 0 7 5
	1 7 0		1 7 0 A

審査請求 未請求 請求項の数 8 O L (全 9 頁)

(21)出願番号 特願2001-353554(P2001-353554)

(22)出願日 平成13年11月19日(2001.11.19)

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72)発明者 牧野 俊朗

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(72)発明者 杉崎 正之

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(74)代理人 100072718

弁理士 古谷 史旺

最終頁に続く

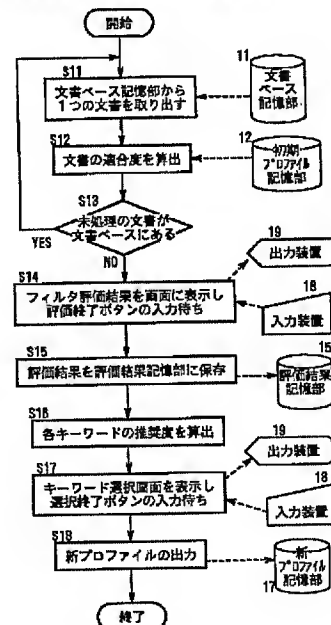
(54)【発明の名称】 プロファイル作成支援方法及び装置並びに記録媒体及びソフトウェア

(57)【要約】

【課題】 本発明は適切なプロファイルの作成を容易にすることが可能なプロファイル作成支援方法及び装置並びに記録媒体及びソフトウェアを提供することを目的とする。

【解決手段】 ユーザの興味を表すプロファイルの作成を支援するためのプロファイル作成支援方法であって、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを初期プロファイルを用いて行い、その結果をユーザに出力し、その結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け、評価結果に基づいて正解の文書及び不正解の文書の内容を解析し、解析の結果を初期プロファイルの情報と比較して、追加候補キーワード及び削除候補キーワードを抽出してユーザに出力し、これらの候補キーワードに対するユーザの選択入力に従って、プロファイルの修正を行うことを特徴とするプロファイル作成支援方法。

実施の形態のプロファイル作成支援装置の動作



【特許請求の範囲】

【請求項1】 ユーザの興味を表すキーワードの集合を含むプロフィールの作成を支援するためのプロフィール作成支援方法であって、

複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロフィールの情報をを用いて行い、

前記フィルタリングの結果をユーザに対して出力し、

前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶し、

前記評価結果に基づいてプロフィールの修正を行うことを特徴とするプロフィール作成支援方法。

【請求項2】 請求項1のプロフィール作成支援方法において、

前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析し、

前記解析の結果を前記初期プロフィールの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出し、

抽出された前記候補リストをユーザに対して出力し、

前記候補リストに対するユーザからの入力に従って、プロフィールの修正を行うことを特徴とするプロフィール作成支援方法。

【請求項3】 ユーザの興味を表すキーワードの集合を含むプロフィールの作成を支援するためのプロフィール作成支援装置であって、

予め用意された初期プロフィールの情報を保持する初期プロフィール記憶手段と、

複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを前記初期プロフィールの情報をを用いて行うフィルタリング手段と、

前記フィルタリングの結果をユーザに対して出力し、その結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶する入出力制御手段と、

前記評価結果に基づいて、プロフィールの修正を行うプロフィール修正手段とを設けたことを特徴とするプロフィール作成支援装置。

【請求項4】 請求項3のプロフィール作成支援装置において、

前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析する文書解析手段と、

前記文書解析手段の解析結果を前記初期プロフィールの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出する候補リスト抽出手段とを更に設け、前記プロフィール修正手段は、前記候補リスト抽出手段の抽出した候補リストをユーザに対して出力するとともに、前記候補リストに

対するユーザからの入力に従ってプロフィールの修正を

行うことを特徴とするプロフィール作成支援装置。

【請求項5】 ユーザの興味を表すキーワードの集合を含むプロフィールの作成を支援するためのコンピュータで実行可能なソフトウェアを記録した記録媒体であって、前記ソフトウェアには、

複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロフィールの情報をを用いて行う手順と、

前記フィルタリングの結果をユーザに対して出力する手順と、

前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付ける手順と、

前記評価結果に基づいてプロフィールの修正を行う手順とを設けたことを特徴とする記録媒体。

【請求項6】 ユーザの興味を表すキーワードの集合を含むプロフィールの作成を支援するためのコンピュータで実行可能なソフトウェアを記録した記録媒体であって、前記ソフトウェアには、

20 複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロフィールの情報をを用いて行う手順と、

前記フィルタリングの結果をユーザに対して出力する手順と、

前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶する手順と、

前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析する手順と、

30 前記解析の結果を前記初期プロフィールの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出する手順と、

抽出された前記候補リストをユーザに対して出力する手順と、

前記候補リストに対するユーザからの入力に従って、プロフィールの修正を行う手順とを設けたことを特徴とする記録媒体。

【請求項7】 ユーザの興味を表すキーワードの集合を含むプロフィールの作成を支援するためのコンピュータで実行可能なソフトウェアであって、

40 複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロフィールの情報をを用いて行う手順と、

前記フィルタリングの結果をユーザに対して出力する手順と、

前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付ける手順と、

50 前記評価結果に基づいてプロフィールの修正を行う手順とを設けたことを特徴とするソフトウェア。

【請求項8】 ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのコンピュータで実行可能なソフトウェアであって、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロファイルの情報を用いて行う手順と、

前記フィルタリングの結果をユーザに対して出力する手順と、

前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶する手順と、

前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析する手順と、

前記解析の結果を前記初期プロファイルの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出する手順と、

抽出された前記候補リストをユーザに対して出力する手順と、

前記候補リストに対するユーザからの入力に従って、プロファイルの修正を行う手順とを設けたことを特徴とするソフトウェア。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、コンピュータを用いて例えば大量の文書の中からユーザの興味に合致する文書を選び出すような情報フィルタに利用されるユーザプロファイル、すなわちユーザの興味を表すキーワードリストのような情報の作成を支援するために用いるプロファイル作成支援方法及び装置並びに記録媒体及びソフトウェアに関する。

【0002】

【従来の技術】近年のインターネットの発達等により、電子化された大量の文書へのアクセスが容易になった。その反面、大量の文書の中から自分の興味に合致する特定の文書を見つけることは困難である。そこで、興味に合致する特定の文書を効率的に見つけるために、キーワードを用いた各種の検索技術や情報フィルタリング技術が用いられている。

【0003】情報フィルタリングを行う場合には、その人の興味を表すプロファイルが必要になる。プロファイルとしては、例えば人手でキーワードを並べたものが用いられる。また、興味の対象となる文書のサンプル集合をコンピュータで解析し、出現単語の統計情報を利用して、単語とその重みを算出しその結果を並べたものを用いる場合もある。

【0004】

【発明が解決しようとする課題】しかしながら、上記のようなプロファイルを作成する場合には、以下のような問題がある。

【0005】まず、人手でキーワードを並べる場合に

は、適切なキーワードを想定することが困難である。適切なキーワードを想定できない場合には、興味対象の文書がうまくフィルタにひっかからなかったり、逆に、想定したキーワードが興味対象でない文書中出现し、結果的にノイズを増やしてしまうという問題がある。

【0006】また、サンプル文書を元にプロファイルをコンピュータで自動作成する場合には、適切な統計処理を行う必要がある。また、統計処理がうまく機能するためには大量のサンプル文書が必要であり、サンプル文書を集めるのが大変である。また、作成されるプロファイルは、サンプル文書の内容に依存するため、適切なプロファイルになっていないことも多い。更に、統計情報に基づいて自動作成されるので、作成されたプロファイルの内容を手作業で適切に修正することは困難である。

【0007】本発明は、適切なプロファイルの作成を容易にすることが可能なプロファイル作成支援方法及び装置並びに記録媒体及びソフトウェアを提供することを目的とする。

【0008】

【課題を解決するための手段】請求項1は、ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのプロファイル作成支援方法であって、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロファイルの情報を用いて行い、前記フィルタリングの結果をユーザに対して出力し、前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶し、前記評価結果に基づいてプロファイルの修正を行うことを特徴とする。

【0009】請求項1においては、初期プロファイルを用いてフィルタリングを行った結果がユーザに提示されるので、プロファイルの作成に利用するサンプル文書をユーザが探す必要はない。すなわち、フィルタリングの結果に対する正解及び不正解を表すユーザからの入力に基づいてプロファイルが修正され、新しいプロファイルが作成される。

【0010】請求項2は、請求項1のプロファイル作成支援方法において、前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析し、前記解析の結果を前記初期プロファイルの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出し、抽出された前記候補リストをユーザに対して出力し、前記候補リストに対するユーザからの入力に従って、プロファイルの修正を行うことを特徴とする。

【0011】請求項2においては、追加候補キーワード及び/又は削除候補キーワードを自動的に作成してユーザに提示するので、ユーザは追加すべきキーワード及び削除すべきキーワードの選択に関する判断を行うだけでよく、正確なプロファイルを簡単に作成できる。また、

実際にフィルタリングを行った際に、興味対象でないドキュメントを拾ってしまうキーワードをユーザは知ることができる。このように、より正確なプロファイルの作成を支援することができる。

【0012】請求項3は、ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのプロファイル作成支援装置であって、予め用意された初期プロファイルの情報を保持する初期プロファイル記憶手段と、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを前記初期プロファイルの情報を用いて行うフィルタリング手段と、前記フィルタリングの結果をユーザに対して出力し、その結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶する入出力制御手段と、前記評価結果に基づいて、プロファイルの修正を行うプロファイル修正手段とを設けたことを特徴とする。

【0013】請求項3のプロファイル作成支援装置を用いることにより、請求項1の方法を実施することができる。請求項4は、請求項3のプロファイル作成支援装置において、前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析する文書解析手段と、前記文書解析手段の解析結果を前記初期プロファイルの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出する候補リスト抽出手段とを更に設け、前記プロファイル修正手段は、前記候補リスト抽出手段の抽出した候補リストをユーザに対して出力するとともに、前記候補リストに対するユーザからの入力に従ってプロファイルの修正を行うことを特徴とする。

【0014】請求項4のプロファイル作成支援装置を用いることにより、請求項2の方法を実施することができる。請求項5は、ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのコンピュータで実行可能なソフトウェアを記録した記録媒体であって、前記ソフトウェアに、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロファイルの情報を用いて行う手順と、前記フィルタリングの結果をユーザに対して出力する手順と、前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付ける手順と、前記評価結果に基づいてプロファイルの修正を行う手順とを設けたことを特徴とする。

【0015】請求項5の記録媒体に記録されたソフトウェアを用いることにより、請求項1の方法を実施することができる。請求項6は、ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのコンピュータで実行可能なソフトウェアを記録した記録媒体であって、前記ソフトウェアに、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロファイルの情報を用いて行う

手順と、前記フィルタリングの結果をユーザに対して出力する手順と、前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶する手順と、前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析する手順と、前記解析の結果を前記初期プロファイルの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出する手順と、抽出された前記候補リストをユーザに対して出力する手順と、前記候補リストに対するユーザからの入力に従って、プロファイルの修正を行う手順とを設けたことを特徴とする。

【0016】請求項6の記録媒体に記録されたソフトウェアを用いることにより、請求項1及び請求項2の方法を実施することができる。請求項7は、ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのコンピュータで実行可能なソフトウェアであって、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロファイルの情報を用いて行う手順と、前記フィルタリングの結果をユーザに対して出力する手順と、前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付ける手順と、前記評価結果に基づいてプロファイルの修正を行う手順とを設けたことを特徴とする。

【0017】請求項7のソフトウェアを用いることにより、請求項1の方法を実施することができる。請求項8は、ユーザの興味を表すキーワードの集合を含むプロファイルの作成を支援するためのコンピュータで実行可能なソフトウェアであって、複数の文書からユーザの興味に合致する特定文書を選び出すフィルタリングを予め用意された初期プロファイルの情報を用いて行う手順と、前記フィルタリングの結果をユーザに対して出力する手順と、前記フィルタリングの結果に対するユーザからの入力を正解及び不正解を表す評価結果の情報として受け付け記憶する手順と、前記評価結果に基づいて、正解の文書及び不正解の文書の内容を解析する手順と、前記解析の結果を前記初期プロファイルの情報と比較して、追加候補キーワード及び削除候補キーワードの少なくとも一方を候補リストとして抽出する手順と、抽出された前記候補リストをユーザに対して出力する手順と、前記候補リストに対するユーザからの入力に従って、プロファイルの修正を行う手順とを設けたことを特徴とする。

【0018】請求項8のソフトウェアを用いることにより、請求項1及び請求項2の方法を実施することができる。

【0019】

【発明の実施の形態】本発明のプロファイル作成支援方法及び装置並びに記録媒体及びソフトウェアの1つの実施の形態について、図1～図6を参照して説明する。こ

の形態は全ての請求項に対応する。

【0020】図1はこの形態のプロファイル作成支援装置の動作を示すフローチャートである。図2はこの形態のプロファイル作成支援装置の構成を示すブロック図である。図3は初期プロファイルの具体例を示す模式図である。図4はフィルタ結果評価画面の表示例を示す正面図である。図5は評価結果記憶部内の評価結果の具体例を示す模式図である。図6は追加削除キーワード選択画面の表示例を示す正面図である。

【0021】この形態では、請求項3の初期プロファイル記憶手段、フィルタリング手段、入出力制御手段及びプロファイル修正手段は、それぞれ初期プロファイル記憶部12、情報フィルタリング処理部13、フィルタ結果評価部14及びプロファイル構築支援部16に対応する。また、請求項4の文書解析手段及び候補リスト抽出手段はステップS16に対応する。

【0022】図2に示すように、この形態のプロファイル作成支援装置は、文書ベース記憶部11、初期プロファイル記憶部12、情報フィルタリング処理部13、フィルタ結果評価部14、評価結果記憶部15、プロファイル構築支援部16、新プロファイル記憶部17、入力装置18及び出力装置19を備えている。

【0023】また、情報フィルタリング処理部13及びプロファイル構築支援部16には、それぞれ関数テーブル20及び21が備わっている。文書ベース記憶部11は、文字列を主体とする様々な文書の情報を大量に保持している。なお、文書ベース記憶部11の代わりにオンラインのデータベースを利用してもよい。

【0024】初期プロファイル記憶部12は、特定のユーザの興味の対象に関する初期プロファイルの情報を保持している。この初期プロファイルは、例えば手作業で予め作成され、初期プロファイル記憶部12に記憶される。この初期プロファイルは、例えば図3に示すように多数のキーワードを並べたキーワードリストである。各キーワードは、特定のユーザの興味の対象と関係する単語の文字列である。

【0025】情報フィルタリング処理部13は、文書ベース記憶部11から文書の情報を1つずつ取り出し、初期プロファイル中の各キーワードの出現頻度を計数し、各文書の適合度を求める。情報フィルタリング処理部13は、全ての文書に対して適合度を求めた後で、予め定めた閾値以上の適合度を持つ各文書の識別子（例えばファイル名）を適合度とともにフィルタ結果評価部14に渡す。

【0026】フィルタ結果評価部14は、情報フィルタリング処理部13から渡された識別子に基づいて文書ベース記憶部11から文書の情報を取り出し、それをフィルタ結果評価画面として出力装置19に出力する。このフィルタ結果評価画面には、図4に示すように、文書毎に正解、不正解を選択するための選択ボックスも表示さ

れる。

【0027】また、フィルタ結果評価部14は、フィルタ結果評価画面に対してユーザから入力される正解、不正解の選択情報に対応する評価結果を評価結果記憶部15に出力する。評価結果記憶部15は、フィルタ結果評価部14から入力された評価結果の情報を保存する。

【0028】プロファイル構築支援部16は、フィルタ結果評価部14に保持されている評価結果に基づいて、正解の文書及び不正解のそれぞれの文書を文書ベース記憶部11から取り出し、各文書の内容を解析する。すなわち、単語の出現頻度を求め、その出現頻度に基づき各単語に推奨度を付与し、追加キーワード候補及び削除キーワード候補を自動的に抽出する。そして、抽出された追加キーワード候補及び削除キーワード候補の情報を、追加削除キーワード選択画面として出力装置19に出力する。

【0029】この追加削除キーワード選択画面には、図6に示すように、各キーワードの選択の有無を指定するための選択ボックスがキーワード毎に表示される。また、プロファイル構築支援部16は、追加削除キーワード選択画面に対して入力装置18によりユーザから入力される選択情報に基づいて新プロファイルの情報を作成する。

【0030】新プロファイルは、初期プロファイルに対してユーザの選択した追加キーワード候補を追加し、削除キーワード候補を削除した結果である。この新プロファイルが新プロファイル記憶部17に出力され記憶される。入力装置18は、ユーザがフィルタ結果に対して正解、不正解の選択入力を行う場合や、追加削除キーワード選択画面に対して追加キーワード候補及び削除キーワード候補の選択入力を行う場合に利用される。

【0031】出力装置19は、ユーザに対してフィルタ結果評価画面や追加削除キーワード選択画面を表示するために利用される。初期プロファイルの具体例は図3に示す通りである。すなわち、ユーザの興味のある文書を特定するための多数のキーワードが並べてある。この例では、1行に1つのキーワードが記述してある。なお、新プロファイルも同様の形式で記述される。

【0032】なお、図2に示す要素の大部分については、専用のハードウェアで構成することもできるし、パソコンなどのプログラム実行により実現することもできる。本発明をプログラムとして実現する場合には、プログラムやデータを例えばCD-ROMのような記録媒体に記録して配布することもできるし、オンラインで転送される情報として配布することもできる。

【0033】フィルタ結果評価部14の制御によって出力装置19に表示されるフィルタ結果評価画面は、例えば図4に示すように構成される。この例では、フィルタ結果評価画面の左側には識別子のリストが存在し、それぞれの識別子の右側に正解及び不正解の選択ボックスが

存在する。また、このフィルタ結果評価画面上でいずれかの識別子をクリックすると、その識別子に該当する文書の内容が、フィルタ結果評価画面の右側に表示される。図4の例では、「10002」の識別子に対応する文書の内容がフィルタ結果評価画面の右側に表示されている。

【0034】ユーザは、フィルタ結果評価画面上で各識別子に対応する文書の内容を見て、正解か不正解かを判定し、その結果に応じて入力装置18を操作し、各選択ボックスの選択を行うことができる。評価結果記憶部15には、例えば図5に示すような評価結果が保持される。この例では、正解及び不正解のそれぞれに対応付けて該当する文書の識別子を並べてある。

【0035】プロファイル構築支援部16の制御によって出力装置19に表示される追加削除キーワード選択画面は、例えば図6に示すように構成される。この例では、追加削除キーワード選択画面の左側に追加候補のキーワードが並べて表示してあり、右側には削除候補のキーワードが並べて表示してある。また、各々のキーワードの近傍には選択ボックスが表示してある。

【0036】ユーザは、追加削除キーワード選択画面上で、入力装置18を用いて各選択ボックスの選択を行うことができる。図2に示すプロファイル作成支援装置の動作は、図1に示す通りである。図1に示す各ステップの処理について以下に説明する。ステップS11では、情報フィルタリング処理部13が文書ベース記憶部11から1つの文書の情報を取り出す。

【0037】ステップS12では、情報フィルタリング処理部13は取り出した文書に対して形態素解析を実行し単語を抽出する。また、初期プロファイル記憶部12に保持される初期プロファイルを参照し、初期プロファイル中のキーワードの出現頻度を計数し、文書の適合度Fを求める。例えば、次式を用いて適合度Fを算出することができる。

【数1】

$$F = \sum_{w_i} \text{freq}(w_i) \cdot \text{wgt}(w_i) \quad \dots (1)$$

w_i : 初期プロファイル中のキーワード

$\text{freq}(w_i)$: 文書中の w_i の出現頻度

$\text{wgt}(w_i)$: w_i の重要度を表す関数

プロファイルは、ユーザが興味を示すキーワードであるので、それが多く含まれている文書はユーザの興味に合致すると考えられる。なお、関数 $\text{wgt}(w_i)$ の値は多くの文書に出現するキーワードに対して小さくなるように決定される。多くの文書に出現する一般的な語は、あまり特徴的ではなく、ユーザが興味を示さない文書中にも出現する確率が高いので、そのような語の重要度を下げることがこの関数の意味するところである。

【0038】よって、前記第(1)式で表される適合度F

が大きいうことは、ユーザが興味を示すより特徴的なキーワードを多く含む文書であり、ユーザの興味により合致していると考えられる。実際には、予め大量のテキスト文書を解析して頻度情報を計数した結果を、関数 $\text{wgt}(w_i)$ の値のテーブルとして関数テーブル20に蓄積してある。従って、情報フィルタリング処理部13は関数テーブル20を参照して関数 $\text{wgt}(w_i)$ の値を取得する。

【0039】図1のステップS13では、文書ベース記憶部11中に未処理の文書が存在するかどうかを識別する。未処理の文書が存在する場合には、ステップS13からS11に戻って処理を繰り返す。全ての文書について処理が終了した場合には、ステップS13で情報フィルタリング処理部13は予め定めた閾値以上の適合度Fを持つ各文書について、その文書識別子及び適合度Fをフィルタ結果評価部14に渡し、次のステップS14に進む。

【0040】この時点で、文書ベース記憶部11の文書に対して初期プロファイルを用いたフィルタリングが完了したことになる。ステップS14では、フィルタ結果評価部14は、情報フィルタリング処理部13から受け取った適合度Fの大きいものから順番に文書の識別子を並べ替え、図4に示すようなフィルタ結果評価画面を作成して出力装置19に出力し、評価結果ボタンの入力を待つ。

【0041】ユーザは、入力装置18を用いてフィルタ結果評価画面上の文書の識別子をクリックすることにより、各文書の内容を表示することができるので、表示される各文書の内容を把握し、正解、不正解の判定をすることができる。また、判定の結果を反映するようにユーザは入力装置18を用いて該当する選択ボックスを選択状態にすることができる。選択が終了してユーザが画面中の「評価終了」ボタンを押下すると、図1のステップS14からS15に進む。

【0042】ステップS15では、フィルタ結果評価部14は、ステップS14で得られた評価結果を例えば図5に示すような形式で評価結果記憶部15に保存する。ステップS16では、プロファイル構築支援部16は評価結果記憶部15の内容である評価結果を参照し、ユーザによって正解と判断された文書及び不正解と判断された文書の識別子を把握する。

【0043】そして、プロファイル構築支援部16は各文書を文書ベース記憶部11から取り出して形態素解析を実行し、各文書から単語を抽出する。更に、抽出された単語が初期プロファイル記憶部12上の初期プロファイルに存在するかどうかを判定し、存在しない場合には追加候補キーワードとしての推奨度を与え、存在する場合には削除候補キーワードとしての推奨度を与える。

【0044】単語 w_i の追加候補キーワードとしての推奨度 D_{add} 及び削除候補キーワードとしての推奨度 D_{del}

はそれぞれ次式により算出することができる。

* * [数2]

$$D_{add} = \left(\frac{\sum d_{ok} \cdot \text{freq}(w_j)}{nd_{ok}} - \frac{\sum d_{ng} \cdot \text{freq}(w_j)}{nd_{ng}} \right) \text{wgt}(w_j) \quad \dots (2)$$

$$D_{del} = \left(\frac{\sum d_{ng} \cdot \text{freq}(w_j)}{nd_{ng}} - \frac{\sum d_{ok} \cdot \text{freq}(w_j)}{nd_{ok}} \right) \text{wgt}(w_j) \quad \dots (3)$$

d_{ok} : 正解と判断された文書

d_{ng} : 不正解と判断された文書

$\sum d_{ok} \cdot \text{freq}(w_j)$: w_j の正解文書集合における出現頻度の和

$\sum d_{ng} \cdot \text{freq}(w_j)$: w_j の不正解文書集合における出現頻度の和

nd_{ok} : 正解文書の数

nd_{ng} : 不正解文書の数

$\text{wgt}(w_j)$: w_j の重要度を表す関数

なお、関数 $\text{wgt}(w_i)$ の値は多くの文書に出現するキーワードに対して小さくなるように決定される。

【0045】実際には、予め大量のテキスト文書を解析して頻度情報を計数した結果を、関数 $\text{wgt}(w_i)$ の値のテーブルとして関数テーブル21に蓄積してある。従って、プロファイル構築支援部16は関数テーブル21を参照して関数 $\text{wgt}(w_i)$ の値を取得することができる。図1のステップS17では、プロファイル構築支援部16はステップS16で得られた追加候補キーワード及び削除候補キーワードのそれぞれを推奨度が大きい順番に並べる。そして、追加候補キーワード及び削除候補キーワードのリストを含む図6に示すような追加削除キーワード選択画面を作成しそれを出力装置19に出力する。更に、プロファイル構築支援部16は画面上の「選択終了」ボタンに対する入力进行を待つ。

【0046】すなわち、正解の文書の特徴付けるキーワードのうち初期プロファイルに存在しないものを追加候補キーワードとして提示し、不正解の文書の特徴付けるキーワードのうち初期プロファイルに存在するものを削除候補キーワードとして提示し、ユーザの選択を促す。ユーザは、入力装置18を用いて入力操作を行うことにより、追加削除キーワード選択画面上の各追加候補キーワード及び削除候補キーワードに対応付けられた選択ボックスに対して選択、非選択の入力を行うとともに「選択終了」ボタンの押下を行うことができる。

【0047】ユーザは、追加削除キーワード選択画面を参照し、追加すべきキーワード及び削除すべきキーワードに関する自分の判断を反映するように各選択ボックスに対する入力操作を行い、更に「選択終了」ボタンを押下する。これによりプロファイル作成支援装置の処理はステップS18に進む。ステップS18では、プロファイル構築支援部16によって新プロファイルが作成さ

れ、この新プロファイルは新プロファイル記憶部17に保存される。

【0048】新プロファイルは、初期プロファイルの内容に基づき、ステップS17で画面に出力された追加候補キーワード及び削除候補キーワードのうち、ユーザによって選択された各キーワードを追加又は削除することによって作成される。この形態では、初期プロファイルに基づいて用意された文書のフィルタリングを行った結果をステップS14でユーザに提示するので、ユーザはプロファイルの作成に必要なサンプル文書を探す必要がない。

【0049】また、フィルタリングの結果に対するユーザの正解、不正解の評価結果に基づいて各文書の解析を行い、正解、不正解にそれぞれ片寄って出現する単語を推奨度に従って追加候補キーワード及び削除候補キーワードとしてステップS17でユーザに提示するので、ユーザは候補の中から追加キーワード及び削除キーワードを判断し選択するだけでよい。これにより、ユーザの興味をより適切に表現するようなプロファイルを簡単に作成することができる。

【0050】

【発明の効果】以上のように、本発明によればユーザの適切なプロファイル構築を支援することができる。また、初期プロファイルに基づいて用意された文書のフィルタリングを行った結果をユーザに提示するので、ユーザはプロファイルの作成に必要なサンプル文書を探す必要がない。

【0051】また、フィルタリングの結果に対するユーザの正解、不正解の評価結果に基づいて各文書の解析を行い、正解、不正解にそれぞれ片寄って出現する単語を推奨度に従って追加候補キーワード及び削除候補キーワードとしてユーザに提示するので、ユーザは候補の中か

ら追加キーワード及び削除キーワードを判断し選択するだけでよい。これにより、ユーザの興味をより適切に表現するようなプロフィールを簡単に作成することができる。

【図面の簡単な説明】

【図1】実施の形態のプロファイル作成支援装置の動作を示すフローチャートである。

【図2】実施の形態のプロファイル作成支援装置の構成を示すブロック図である。

【図3】初期プロフィールの具体例を示す模式図である。

【図4】フィルタ結果評価画面の表示例を示す正面図である。

【図5】評価結果記憶部内の評価結果の具体例を示す模式図である。

*

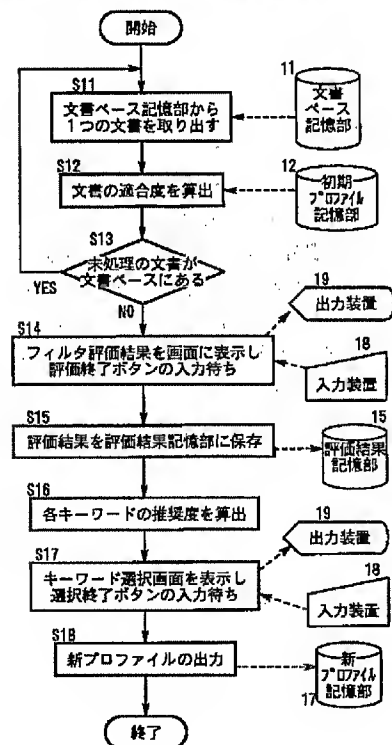
*【図6】追加削除キーワード選択画面の表示例を示す正面図である。

【符号の説明】

- 11 文書ベース記憶部
- 12 初期プロフィール記憶部
- 13 情報フィルタリング処理部
- 14 フィルタ結果評価部
- 15 評価結果記憶部
- 16 プロファイル構築支援部
- 17 新プロフィール記憶部
- 18 入力装置
- 19 出力装置
- 20 関数テーブル
- 21 関数テーブル

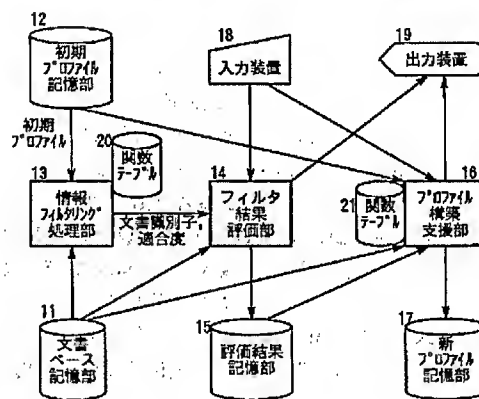
【図1】

実施の形態のプロファイル作成支援装置の動作



【図2】

実施の形態のプロファイル作成支援装置の構成



【図3】

初期プロフィールの具体例

技術
熊鷹堂
修行
職人
伝統
伝統的工芸品
南部鉄器
博多人形
輪島塗
...

【図5】

評価結果記憶部内の評価結果の具体例

正解	不正解
10002	10020
10007	10110
10034	.
10073	.
10143	.
.	.
.	.

【図4】

フィルタ結果評価画面の表示例

フィルタ結果評価画面			
評価終了			
識別子	正解	不正解	
10002	<input checked="" type="checkbox"/>	<input type="checkbox"/>	文書10002 日本の伝統工芸品で… …輪島塗が…
10007	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
10020	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
10034	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
10073	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
10110	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
10143	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
⋮	⋮	⋮	
⋮	⋮	⋮	
⋮	⋮	⋮	

【図6】

追加削除キーワード選択画面の表示例

追加削除キーワード選択画面			
選択終了			
追加候補キーワード*	選択	削除候補キーワード*	選択
通産省	<input type="checkbox"/>	技術	<input checked="" type="checkbox"/>
指定	<input type="checkbox"/>	職人	<input checked="" type="checkbox"/>
輪	<input type="checkbox"/>	修行	<input checked="" type="checkbox"/>
伝統工芸品産業	<input checked="" type="checkbox"/>	伝統	<input checked="" type="checkbox"/>
伝統工芸品	<input checked="" type="checkbox"/>		
原材料	<input type="checkbox"/>		
漆器工房	<input checked="" type="checkbox"/>		
⋮	⋮		
⋮	⋮		
⋮	⋮		

フロントページの続き

(72)発明者 稲垣 博人

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

Fターム(参考) 5B075 ND02 PR08